



MIGRATION OF RELATIONAL DATABASE TO MICROSOFT AZURE CLOUD

Omkar Dhamale | Rushikesh Birajdar | Harshal Wadkar | Rishabh Singh | S. G. Malvi | S. G. Patil

ABSTRACT

The Windows Azure is a Cloud Computing Platform that provides its users to store large amount of data with efficient storage. The infrastructure is created by the Microsoft Corporation and it manages the applications and its products through a global network known as Microsoft azure. The one of the important advantage of using azure is that the data which is stored in azure is present for any duration of time and it can be easily accessible to user at any time and any place. The Structured Query Language i.e. SQL is having some limitations as data is not retrieved efficiently. Also flexibility is also one of the issues with SQL. So in this paper, the purpose is to migrating the SQL queries to Azure cloud which uses NoSQL structure for storing data. The data which is migrated is also stored in normalized form in azure cloud. Based on some experimental results, our mechanism of migration will help the current application system which uses SQL to improve its performance and maintains security.

KEYWORDS: Cloud Storage, distributed storage systems, Windows Azure, SQL, Not Only SQL (NoSQL), data migration.

Introduction:

Windows Azure Cloud is storage platform which is in produced by Microsoft Corporation. This is used in many application areas which includes customer managing systems, retailers of shops, social networking sites and so on. The advantage of using Azure is that it gives customers very large amount of space to store their data and they also manages that data.

As we are using NoSQL, it provides various advantages over the SQL such as it does not provide tabular relationship of database rather than this they provides document which contains Key-Value pair. The data is efficiently accessed and also it represented in normalized form. The SQL does not support the JOIN operation so, when we are migrating it to NoSQL it supports the join operation and also performance gets increased.

As many industries uses SQL only it is efficient to them to use the Microsoft azure cloud storage which can give high performance.

Literature Survey:

The conversion of schema model to SQL database to NoSQL is presented by the Gansen Zhao, Qiaoying Lin, Libo Li, Zijing Li in their research paper. This paper focuses on schema conversion of SQL Databases to the NoSQL. This paper gives the comparison of query performance between Nested solution and Non-Nest solution. As we all known, if the SQL query statement involves multi-table join, the query need to access more than one table in original relational database. The advantage of this paper is the correctness of the conversion algorithm has been conducted a rigorous proof. They also stated that the future work about the subject is to minimize the spatial redundancy [1].

Another paper is about the schema demoralization and migration by using the study of the various content management system. Content Management System are able let people who have no technical skill to manage website, rapidly create, edit or publish online content. so the paper purposes that as the exploring growth huge amount of data the well-structured characteristics of SQL may limits to handle scalability of the data so the use of migration from SQL to NoSQL.

In this paper the flow about the conversion is described with various steps. Advantage of this paper is that the migration useful for various Content Management Systems described thoroughly [2].

“Windows Azure Storage: A Highly Available Cloud Storage Service with Strong Consistency”, is the another paper by Brad Calder, Ju Wang, Aaron Ogus, Niranjana Nilakantan, Arild Skjolsvold, Sam McKelvie.

The paper focuses on the Windows Azure Storage and its various description. Currently, WAS storage comes in the form of Blobs (Files), Tables (structured storage), and Queues (message delivery). In this paper, they describe the WAS architecture, global namespace, and data model, as well as its resource provisioning, load balancing, and replication systems.

This paper concludes that The Windows Azure Storage platform implements essential services for developers of cloud based solutions. The combination of strong consistency, global partitioned namespace, and disaster recovery has been important customer features in Window Azure Storage (WAS's) multitenancy environment [3].

Proposed system:

In our system we are accepting existing relational database, identify table joins and move data to azure tables in normalized form. So that data redundancy can be eliminated. Provide a command line utility which can accept the old queries (SQL) and give same results with data in Azure cloud.

E.g. consider a simplified scenario that there are following tables in a business's database-

Customer (cust_id, name, address)

Item (item_id, description, price)

Purchases (purchase_id, cust_id, item_id, date, amount)

When the Purchases table is moved to Azure Table in normalized way it would probably look like following

Normalized Purchase (purchase_id, cust_id, item_i, customer_name, customer_address, item_description, item_price, purchases_date, purchases_amount)

There would be a single normalized azure table for all of the purchases containing each record.

System Architecture-

System consist of four main components Data Migrator, Query Parser, Normalization, Table Join. System take input as Relational Database process it & convert it into azure table storage format i.e. is the NOSQL.

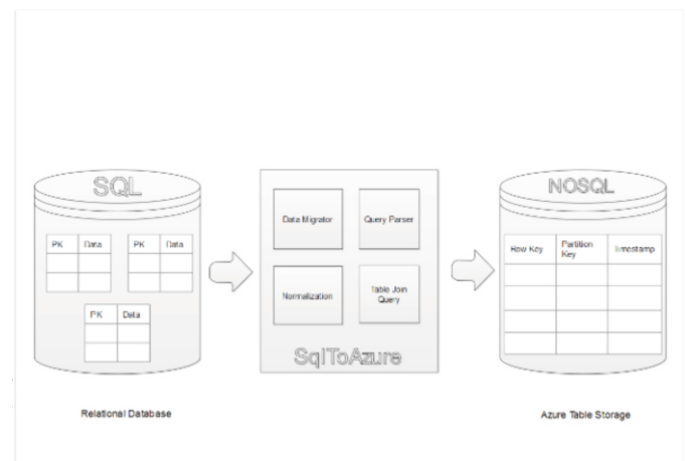


Fig. System Architecture

There are four modules in system-

- 1) **Data migrator**- Data migrator is use to move data from SQL to the azure table storage. Data which is migrated from SQL to Azure Table storage is in the normalization form. Data in the SQL database is in the structure format and when data is migrated from SQL to Azure table storage it is in the structured NoSQL data in the cloud.
- 2) **Query Parser**- To accept existing SQL queries and dynamically translate those to azure data fetch logic. Query Parser module Provide a command line utility which can accept the old SQL queries and give same results with data in Azure cloud.
- 3) **Normalization**- Normalization module is use to reduce data redundancy and improve data integrity. Normalization means arranging attributes in tables based on dependencies between attributes, ensuring that the dependencies are properly enforced by database integrity constraints.

List of Normal Forms

- UNF - "Unnormalized Form"
- 1NF - First Normal Form
- 2NF - Second Normal Form
- 3NF - Third Normal Form
- EKNF - Elementary Key Normal Form
- BCNF - Boyce-Codd Normal Form
- 4NF - Fourth Normal Form
- ETNF - Essential Tuple Normal Form
- 5NF - Fifth Normal Form
- 6NF - Sixth Normal Form
- DKNF - Domain/Key Normal Form

- 4) **Table join**- Table join module identify the table joins in relational database. Tables and relations in entities to store entities in NOSQL databases.

Conclusion:

The Windows Azure Storage platform provides essential services for developers of cloud based applications. The Windows Azure Storage platform provides strong consistency, global partitioned namespace, and disaster recovery for large scale applications. In this article we propose sql to azure intermediate system which will accept the relational database and migrate those data to azure table storage which is nosql for cloud based applications in normalized form.

REFERENCES

- [1] Gansen Zhao Qiaoying Lin Libo Li Zijiang Li: Schema Conversion Model of SQL Database to NoSQL, Ninth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (2014).
- [2] Chao-Hsien Lee, Yu-Lin Zheng: SQL-to-NoSQL Schema Denormalization and Migration: A Study on Content Management Systems, IEEE International Conference on Systems (2015).
- [3] Brad Calder, Ju Wang, Aaron Ogus, Niranjan Nilakantan, Arild Skjolsvold, Sam McKelvie, Yikang Xu, Shashwat Srivastav, Jiesheng Wu, Huseyin Simitci, Jaidev Haridas, Chakravarthy Uddaraju, Hemal Khatri, Andrew Edwards, Vaman Bedekar, Shane Mainali, Rafay Abbasi, Arpit Agarwal, Mian Fahimul Haq, Muhammad Ikramul Haq, Deepali Bhardwaj, Sowmya Dayanand, Anitha Adusumilli, Marvin McNett, Sriram Sankaran, Kavitha Manivannan, Leonidas Rigas : Windows Azure Storage: A Highly Available Cloud Storage Service with Strong Consistency, ACM Transactions on Computer Systems, (2011).
- [4] Dongzhao Liang, Yunzhen Lin, Guiguang Ding : Mid-model Design Used in Model Transition and Data Migration between Relational Databases and NoSQL Databases, IEEE International Conference, (2015).
- [5] Erik Meijer and Gavin Bierman, Microsoft Research: A co-Relational Model of Data for Large Shared Data Banks, ACM Transactions, (2011).
- [6] Eric A. Brewer, "Towards Robust Distributed Systems. (Invited Talk)," in Principles of Distributed Computing, Portland, Oregon, (2000).
- [7] F. Chang et al., "Bigtable: A Distributed Storage System for Structured Data," in OSDI, 2006.
- [8] S. Lombardo, E. Di Nitto, and D. Ardagna, "Issues in Handling Complex Data Structures with NoSQL Databases," Proceedings of the 14th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), 2012.